

Using Differential Privacy to Protect Earnings and Employment Statistics

Discussant Comments

John Sandoval

National Advisory Committee

U.S. Census Bureau

2022 Fall Meeting

October 27, 2022

Two important population cohorts that are top of mind



Pandemic's negative impact on education is just starting to show in national indicators



EducationWeek®

STUDENT ACHIEVEMENT

Two Decades of Progress, Nearly Gone: National Math, Reading Scores Hit Historic Lows



By [Sarah D. Sparks](#) — October 24, 2022 ⌚ 8 min read

Same impact at play in Veteran transition to Civilian life

What We've Learned Serving Military-Connected Communities During COVID-19



Post secondary education industry fuels 3rd party demand and supply of micro data underscoring the need for Disclosure Avoidance

Home / Education / Best Colleges / 10 Majors With Highest Starti...

10 College Majors With the Highest Starting Salaries

Median starting salaries for these majors range from \$71,000 to \$80,300, Payscale data shows.

By [Jackson Nimesheim](#) | July 19, 2022, at 4:42 p.m.

FORTUNE | EDUCATION

ARTICLES CAREER GUIDES RANKINGS ▾ MORE ▾

ARTICLES » THESE ARE THE HIGHEST PAYING COLLEGE MAJORS OF 2022

These are the highest paying college majors of 2022

BY MEGHAN MALAS

April 26, 2022, 9:41 AM



Lack of knowledge and understanding of Differential Privacy among general public highlights the need to de-mystify DP

Reporting on Differential Privacy highlights tradeoff between accuracy and privacy



New system to protect census data may compromise accuracy, some experts say

By Tara Behrampour and Marissa J. Lang
June 1, 2021 at 7:09 p.m. EDT

ARIZONA

The census has a new process to protect your privacy. It also risks a less accurate count

 **Geoff Hing**
Arizona Republic
Published 8:30 a.m. MT Aug. 10, 2021 | Updated 3:45 p.m. MT Aug. 10, 2021

Los Angeles Times

WORLD & NATION

Judges hear arguments over census privacy tool

Bloomberg CityLab

Sign In

Data Scientists Square Off Over Trust and Privacy in 2020 Census

As the Census Bureau releases redistricting data, researchers are split over a security measure called differential privacy that can blur population figures.

The New York Times

 **TheUpshot**

To Reduce Privacy Risks, the Census Plans to Report Less Accurate Data

Guaranteeing people's confidentiality has become more of a challenge, but some scholars worry that the new system will impede research.

Some media outlets explicitly showcase the potential negative impacts of Differential Privacy




The 2020 U.S. census is about to drop. But a new government privacy measure could mean chunks of the data are useless

 **Nanci Sandoz**
Aug. 10, 2021 | Updated: Aug. 11, 2021 6:46 a.m.



Urban Wire :: Research Methods and Data Analytics

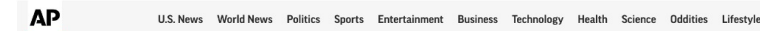
The blog of the Urban Institute

 Will the Census's Data Privacy Efforts Erase Rural America?
March 4, 2020



NEWS | RACIAL JUSTICE

New Census Algorithm Grossly Miscalculates People of Color in the Name of "Privacy"

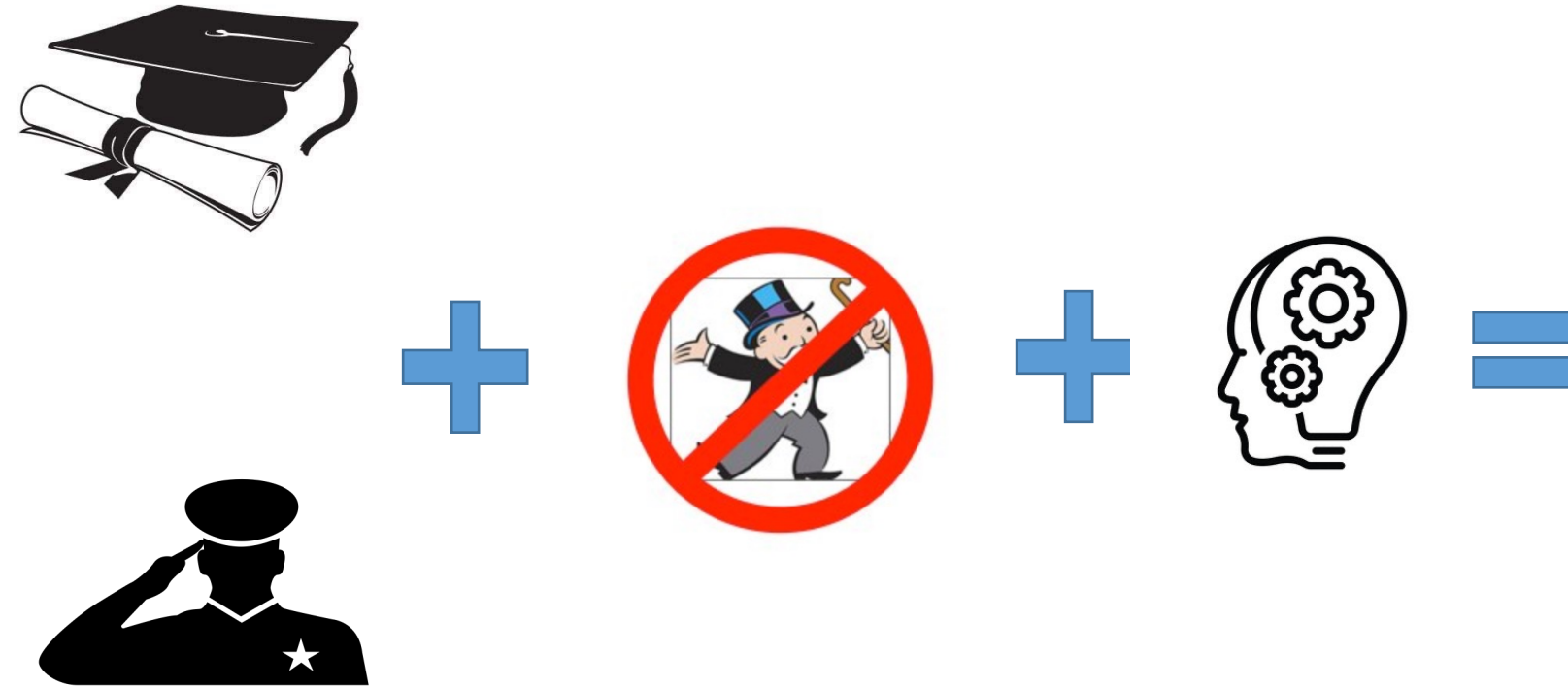


People, homes vanish due to 2020 census' new privacy method

By MIKE SCHNEIDER October 31, 2021



Earning and Employment Statistics may be a unique use case well suited to showcase Differential Privacy in action



Questions / Considerations / Areas to Explore

- Referencing slide 12, can you provide examples of cell sizes of 3—50, 50-80, 80-100 where the mean absolute error exceeds 1000 and percentile errors on the ends of the spectrum show upwards of 4,000 error?

Is it possible these cells could intersect with populations of historically undercounted groups?

Are there instances of intersections with historically undercounted groups the in the PSEO tables?

What about the equivalent table for Veteran Outcomes?

Questions / Considerations / Areas to Explore

- Where was the other 1.5 of privacy budget spent? Can you show your work? Is this an intuitive example that can be used to illustrate the concept of “privacy budget” and the role it plays in Differential Privacy?
- What specific code is being made available to the public on Github? What are the potential use cases for this code? Are there any concerns regarding security?
- On slide 10, can you explain the diversions of the dotted protected CDF from the True CDF? Why are the discrepancies clustered from 50 and 100 on the Earnings axis?
- Can you share the values of the relative accuracy equation from slide 16 for Differential Privacy, evenly spaced bins and smooth sensitivity? Is DP protection method significantly more accurate than the other two?
- On slide 17, is there a line that represents the Differential Privacy method, as compared to evenly spaced bins and smooth sensitivity?